DOCUMENT RESUME

ED 030 410

JC 690 173

By-Connor, Aikin
Is It Really a Better Technique?
California Univ., Los Angeles. ERIC Clearinghouse for Junior Coll. Information.
Report No-Topical Pap-6
Pub Date Mar 69
Note-21p.
EDRS Price MF-$0.25 HC-$1.15
Descriptors-*Junior Colleges, *Program Evaluation, *Statistics

This report presents a "cookbook" approach to assist any teacher or researcher not familiar with statistics in comparing performance differences between two classes using group medians. Illustrations are provided for comparing group medians in hypothetical situations in several disciplines (English, math, philosophy, political science). (JC)

EDUCATIONAL RESOURCES INFORMATION CENTER

ERIC

Clearinghouse for Junior College Information

UNIVERSITY OF CALIFORNIA, LOS ANGELES

TOPICAL PAPERS

IS IT REALLY A BETTER TECHNIQUE?

IS IT REALLY A BETTER TECHNIQUE?

Aikin Connor
Assistant Research Educationist

ERIC Clearinghouse for Junior College Information
Graduate School of Education and the University Library
University of California
Los Angeles, 90024

Topical Paper No. 6
March 1969

This Topical Paper was prepared pursuant to a contract with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official Office of Education position or policy.

TOPICAL PAPERS

1) A Developmental Research Plan for Junior College Remedial Education, July 1968.

2) A Developmental Research Plan for Junior College Remedial Education
   Number 2: Attitude Assessment, November 1968.

3) Student Activism and the Junior College Administrator: Judicial Guidelines, December 1968.

4) Students As Teachers, January 1969.

5) Is Anyone Learning to Write?, February 1969.

6) Is It Really a Better Technique?, March 1969.

INTRODUCTION

This Topical Paper is sixth in a series designed to stimulate research in the junior college. Each presents a model that may be followed by instructors, administrators, or researchers who wish to study effects of their efforts. These Topical Papers are available from the Clearinghouse on request.

Each user of one of these designs is invited to send his results to the Clearinghouse, using either the form provided on page 17 or his own mode of reporting. The reports will be collated and the findings disseminated widely.

The author is a member of the Clearinghouse staff. His design was reviewed by Maurice Smith of Golden West College and Ruth Cline of Los Angeles Valley College.

Arthur M. Cohen
Principal Investigator and Director
ERIC Clearinghouse for Junior
College Information

IS IT REALLY A BETTER TECHNIQUE?

A junior college English instructor thinks his more mature evening students are performing better than his day students. A philosophy professor feels that reprimanding students about performance on exams does more harm than good. A political science instructor calls on the library staff to help her increase student use of current history materials. A math professor notices that classes seem to be requiring less time to cover the same material. What do all these teachers have in common? All are interested in student learning and all are on the verge of formulating explanations that can be tested for accuracy. Suppose, however, none of them is well-versed in educational experimental design or statistics -- must he rely on subjective evaluations and accept his own hunches? The purpose of this paper is to present an easy-to-use plan that offers the power and objectivity of statistics without the complexities.

The basic plan for all research is to record observations. The biologist records what he sees through his microscope; the chemist, what he observes in his test tubes; the educator, what he notes in his classroom. All observe, record, and analyze, but, to have order in their findings, they must first have order in their observations. They carefully design their research so that their observations are focussed, usually on one "variable." By controlling the situation in which this variable operates, the researcher can feel some confidence in interpreting his findings and making generalizations.

The educational researcher may have the most difficult problem of all, since he can rarely control the teaching-learning situation. A simple and, for

the most part, effective plan to achieve the effect of control without actually exerting control is the "control group vs. experimental group" design. In this design, the researcher utilizes two highly similar groups -- one in which the experimental variable is active and one in which it is not. Typical experimental variables in educational research are innovative classroom procedures, experimental testbooks, and new groupings of students. In research parlance, the experimental group is said to have received a "treatment" and the problem of the researcher is to evaluate the effect of the treatment. He records his observations of the groups in terms of, naturally, observable behavior. To impose objectivity on his observations he usually applies some sort of standard measure to the behavior of the two groups. Frequently, this measure is an examination. If the experimental or treated group performs better than the control or untreated group, the experiment or innovation is said to be successful.

The junior college English instructor who hypothesizes that his more mature evening students are better than his day students can observe how the variable of maturity affects students' behavior on examinations. The philosophy professor who thinks that scolding students about grades is harmful can test his hypothesis with verbal chastisement as the experimental variable. The political science instructor can check the effect of her efforts to get students to use the library by comparing those who had received the "treatment" with those who had not. The math professor may have a hunch that the new math program in the state's public schools actually is "paying off" in terms of student performance in college. He therefore compares this year's students with students from previous years. The instructor can test these hunches by means of group

comparisons if he cares to do so.

Having selected comparable groups and subjected one to the proposed innovation (or, as in some of the cases above, having selected groups comparable except in the variable under study), the researcher must make the comparisons. In educational research the relevant measure to compare is some sort of academic performance, usually on a test. In the previous examples, the English instructor may give the same final exam to each class, the philosophy professor also may use scores from the students' tests, the political science instructor may compare numbers of visits to the library, and the math professor may compare results on a series of departmental exams.

If the measured performance of the groups being compared is vastly different, the researcher probably can feel that the variable he is studying is indeed effective. If the performance of the two groups is not really very different, however, he may have some doubts. One way to resolve his doubts is to repeat the experiment or research again and again with different students. If the results are nearly always similar, his doubts may be resolved, since such repetition, or replication, is frequently impractical or impossible, the researcher turns to statistical procedures that offer substantially the same assurance of the reliability of his observations without the need to repeat the research.

The question statistics will answer is "What are the chances of getting similar results if the experiment is repeated?" Usually, educational researchers are content to accept results whose probabilities of being repeated are 95 out of 100. This represents the so-called ".05 level of significance."

Technically, statistical tests are designed to test the hypothesis that two groups actually represent subsets or different parts of some larger group. Relating this kind of test to the comparison of experimental and control groups, statistical tests will indicate the probability that the two groups are really only subsets of a larger group -- that, even though one has had a "treatment" or does differ on some variable, the measure on which they are compared does not reflect the difference significantly. If the probability of being from the same larger group is 5 per cent or less, the treatment or variable being researched is assumed to be responsible.

The researcher then gathers his data, that is, his measured comparisons, and proceeds to analyze his findings statistically. The statistical test he chooses must depend on the specific information he seeks and the specific assumptions regarding his data that underly various statistical tests.

In the simple design of group comparisons, an easy method of determining the reliability of the results is the following adaptation of a general statistical procedure called the Median Test. Since it ignores the size of differences between scores on the comparison measure, the Median Test is not the most powerful of statistical tests. However, in many educational research problems the size of differences on a comparison measure may be slightly inaccurate and a so-called "powerful" test might, in effect, exaggerate this failing. There are two important factors favoring the Median Test for group comparisons -- it is easy to compute and requires only two basic assumptions. One assumption is that both groups are compared by the same measure; the second, that the two groups are, in fact, separate groups, not the same group measured before and

after a treatment. In the following presentation, the Median Test has been restructured, so to speak, to facilitate computation and to avoid presenting unfamiliar concepts and procedures.

Step 1. Find the median for both groups combined. (Directions for finding the median are given in Appendix A.)

Step 2. Find the difference between the number of scores in the experimental group above the combined median and the number below. If there are fewer above the median than below, there is no need to continue -- the treatment was not successful.

Step 3. Repeat Step 2 for the control group.

Step 4. Find the average difference, using both groups. (Number from Step 2 plus number from Step 3, divided by 2.)

Step 5. If the number of scores above or below the median in either group is fewer than 10, subtract 1 from the average difference, as determined in Step 4.

Step 6. Square the difference found in Step 5 (or Step 4, if Step 5 is skipped).

Step 7. Divide the number from Step 6 by the number of scores in the experimental group above the median plus the number below.

Step 8. Divide the number in the experimental group (those above the median plus those below) by the number in the control group above the median plus the number below.

Step 9. Multiply the number from Step 7 by the number from Step 8 plus 1.

Step 10. If the number from Step 9 is greater than 2.71, the two groups may be presumed to be significantly different -- the treatment was a success.*

*The number 2.71 is the size of Chi-square necessary to reject the null hypothesis in a one-tailed test, with one degree of freedom, at the .05 level of confidence.

A special case where this test may be used without a specific control group is in comparing the experimental group's performance on a standardized test with the published norms for the test. In this special case, Step 1 is given as the 50th percentile taken from the norms. Follow Steps 2, 5, 6 and 7. If the figure from Step 7 is larger than 2.71, the experimental group may be assumed to be significantly different.

For some real-life examples of this procedure, let us return to those teachers described earlier. In each case, the necessary date (i.e., test scores, number in each group, etc.) will be given. So that the reader, if he chooses, may practice using the procedure, answers for each step will be given in the margin, and may be covered.

Case No. 1

A college English instructor had a feeling that his evening English I class "went better" than his daytime class, he felt the students responded better and that he taught better. His hunch was that the evening students, being more mature in age, were more mature in their general understanding. Knowing, however, that appearances can be deceptive and that they might not really be so different from the less mature daytime students, he gave each class a test covering the objectives he had projected at that point. On the basis of this test, he compared the two groups, considering the evening class as the experimental group (i.e., the class which had measurably greater maturity in age), with the simplified adaptation of the Median Test. The class scores are given on the facing page.

Evening Class (Experimental)	Daytime Class (Control)
95	98
93	92
90	88
87	88
86	85
86	82
86	81
85	80
85	79
84	76
80	76
79	76
78	75
77	
76	
76	

Step 1.	Median for two groups combined.	(84)
Step 2.	Difference in experimental group scores between those above and those below the combined median.	(3)
Step 3.	Difference in control group scores between those above and those below the combined median.	(3)
Step 4.	Average difference between the two groups.	(3)
Step 5.	Since number of scores below the median is less than 10, subtract 1 from the average difference.	(2)
Step 6.	Square number from Step 5.	(4)
Step 7.	Divide number from Step 6 by 15 (total in evening class).	($4 \div 15 = .27$)
Step 8.	Divide number in evening class by number in daytime class.	($15 \div 13 = 1.15$)
Step 9.	Multiply 1 plus number from Step 8, by number from Step 7.	($2.15 \times .27 = .57$)
Step 10.	As number from Step 9 is less than 2.71, the difference between the classes is not significant.	

Case No. 2 *

A junior college philosophy professor had heard his colleagues talk of the efficacy of tongue-lashing on student's grades, but had not heard of any empirical data -- only opinions. One semester, when he had two sections of an introductory course in philosophy that seemed comparable in ability, he planned to gather the empirical data he thought would refute his colleagues' opinions. For one class, he followed his usual methods of teaching and grading. For the other class, he taught the same way but, instead of grading as usual, kept two sets of grades. One set was the "real" grade, which he recorded; the second set of grades appeared on the students' exams and each grade was systematically reduced two letter grades. Besides giving the apparent low grades in this class, he berated the students after each test to "shape up or ship out." His final comparison of the groups was the final exam, which he graded in his customary method for both classes. The scores are given on facing page.

* This example was suggested by a study made by Dr. Lawrence A. Wenzel, Chico State College, while he was teaching philosophy at Yuba College.

Control Class	Experimental Class
98	95
95	90
94	87
92	85
90	80
90	78
88	76
88	74
87	74
86	72
84	70
80	68
80	65
78	60
75	58
71	57
68	50

Step 1.	Median of combined groups.	(79.83)
Step 2.	Difference in experimental class between scores above median and below.	(7)
Step 3.	Difference in control class between scores above and below the median.	(9)*
Step 4.	Average difference between two classes.	(8)
Step 5.	Since number of scores above the median was less than 10, subtract 1 from average difference.	(7)
Step 6.	Square the difference.	(49)
Step 7.	Divide by total in the control class.	($49 \div 17 = 2.88$)
Step 8.	Divide number in experimental class by number in control class.	($17 \div 17 = 1$)
Step 9.	Multiply 1 plus number from Step 8, by number from Step 7.	($2 \times 2.88 = 5.76$)
Step 10.	As number from Step 9 is greater than 2.71, he concludes that his hunch was right -- berating students is harmful to their academic performance.	

---

* Remember -- his hypothesis was that the experimental group would perform <u>less</u> well than the control group.

Case No. 3

A junior college political science instructor had felt for some time that her students didn't take proper advantage of the library facilities. She discussed the problem with the library staff and together they developed a short instructional program about the library. Since the program utilized class time and involved the library staff, the instructor thought it wise to evaluate its benefits before she used it full-scale. Because the program had informational content, it would have been possible for her to test the students in terms of their acquisition of such information. However, as a political science instructor, she was really interested only in whether or not the program resulted in more student use of the library. She therefore took as a criterion measure the number of books checked out during a semester. She compared the class that had been given the instructional program with another to whom she taught the same course. The number of books each student checked out during the semester is given on the opposite page.

Control Class	Experimental Class
11	10
6	9
4	8
3	8
3	7
3	6
3	6
3	5
2	3
2	3
1	2

Step 1.	Median of two classes combined.	(3.5)
Step 2.	Difference in experimental class scores above median and below.	(5)
Step 3.	Difference in control class above and below median.	(5)
Step 4.	Average difference between the two classes.	(5)
Step 5.	Since both above median and below median categories are fewer than 10, subtract 1.	(4)
Step 6.	Square number from Step 5.	(16)
Step 7.	Divide by total in experimental class.	$(16 \div 11 = 1.45)$
Step 8.	Divide number in experimental class by number in control class.	$(11 \div 11 = 1)$
Step 9.	Multiply 1 plus number from Step 8, by number from Step 7.	$(2 \times 1.45 = 2.90)$
Step 10.	Since number from Step 9 is more than 2.71, the instructional program was considered successful.	

Case No. 4

A math professor noticed that this year's class was substantially farther ahead, in terms of concept understanding, than classes in previous years. If this observation were really true, he thought it might be due to the effect of the new math program in the state, since this year's students had all been exposed to it through most of their public school years. Fortunately for him, the math department had given departmental exams each year and had kept a record of the results for the previous ten years. Thus he was able to compare his current students' performance with that of students from the past ten years. Because the departmental exam was, in effect, a standardized test and had norms developed for it, he was able to use the 50th percentile in place of the median. The scores of his current students and the 50th percentile of the norm group are given opposite.

Current Math Class Scores

50th Percentile = 76

98
96
96
95
95
95
93
90
87
86
85
85
84
80
78
75
69
65
60

Step 1. 50th percentile given as 76.

Step 2. Difference between scores above and below. (11)

Step 5. Since only 4 students scored below the 50th percentile, subtract 1 from the difference. (10)

Step 6. Square number from Step 5. (100)

Step 7. Divide by total in class (100 ÷ 19 = 5.26)

Step 8. Since the number from Step 7 is larger than 2.71 he concluded that this year's class was significantly more proficient in math.

As an additional aid to the would-be researcher, Table I, Appendix A, gives the number of cases in the experimental group above the median necessary to achieve a significant difference. Directions for its use are given beneath the table. Note that if the number of scores of the two groups combined above the median is different from the number below the median by more than one, Table I can not be used. In this relatively rare instance, it will be necessary to make the computations as outlined on page 5. Computation will also be necessary if either group numbers more than 20.

Appendix A

The median is a point (on the same scale used to measure the group) that divides the upper half of the scores from the lower half. Each score is assumed to be the midpoint of a "score interval." For example, a score of 8 is the midpoint of the score interval 7.5-8.5, a score of 123 is the midpoint of the score interval 122.5-123.5, etc.

To find the median if the total number in the group is odd:

1. Arrange scores in order, low to high.
2. Subtract 1 from total number and divide the remainder by 2.
3. From the lowest score, count until the number from Step 2 is reached.
4. Median is the next score above.

Example: Scores 1, 3, 5, 9, 11 N=5

$5 - 1 = 4$; $4/2 = 2$; score of 3 is 2nd score, score of 5 (the next score above) is the median.

To find the median if the total number in the group is even:

1. Arrange scores in order, low to high.
2. From the lowest score, count upward until half the scores are counted.
3. The median will be a point halfway into the interval between the highest point of the score interval represented by the top score in the lower half, and the lowest point of the score interval represented by the lowest score in the upper half.

   Example: Scores 1, 2, 3, 5, 8, 8, 10, 11 N=8

   Median is located halfway between 5.5 (the upper limit of the highest score interval in the lower half) and 7.5 (the lower limit of the lowest score interval in the upper half). Thus, the median is $(5.5 + 7.5)/2$ or 6.5.

Example: Scores 1, 2, 5, 5, 5, 8, 10, 11 N=8

Since the interval between the highest score in the lower half and the lowest score in the upper half is occupied by the three scores of 5, the median is located after the 2nd score of 5 and before the 3rd. Thus, since the lower limit of the score interval represented by 5 is 4.5, the median is 4.5 +.67 (2/3rds of the way through the score interval) or 5.17.

TABLE I.

	Number in Control Group										
Number in Experimental Group	10	11	12	13	14	15	16	17	18	19	20
10	8	8	8	8	8	8	8	8	8	8	8
11	8	8	9	9	9	9	9	9	9	9	9
12	9	9	9	9	9	9	9	9	9	9	9
13	10	10	10	10	10	10	10	10	10	10	10
14	10	10	10	10	10	10	10	10	10	11	10
15	11	11	11	11	11	11	11	11	11	11	11
16	11	11	11	11	11	11	11	11	11	12	11
17	12	12	12	12	12	12	12	12	12	12	12
18	12	12	12	12	12	12	12	12	12	13	13
19	13	13	13	13	13	13	13	13	13	13	13
20	13	13	13	13	13	13	13	13	13	14	14

Find the total in the experimental group in the vertical column; follow across to the number in the control group. Number in the intersection is the number required in the experimental group above the median for significant difference. If the difference between the total number of scores above the median (experimental group plus the control group) and the total number below the median is more than one, this table is not accurate.

Form For Reporting Research

Researcher's Name

Name of College

Subject Matter Area of the Research

Location of College

In the space below, describe the research. The description should include: (1) hypothesis (or "hunch"), (2) number of students involved in each group, (3) statement of procedure, (4) results.

Please mail form to:
ERIC Clearinghouse for Junior College Information
96 Powell Library, UCLA
Los Angeles, California 90024